REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Service, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From-to)*
06-MAR-2009	Final Technical Report	04-May-06 to 30 SEP-07

4. TITLE AND SUBTITLE
New Theory and Algorithms for Compressive Sensing

5a. CONTRACT NUMBER

5b. GRANT NUMBER
N00014-06-1-0769

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Baraniuk, Richard G.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Rice University
6100 Main St. MS 16
Houston TX 77005

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Reza Malek-Madani
Office of Naval Research, Code 311
875 N Randolph St.
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSORING/MONITORING AGENCY REPORT NUMBER

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

In this project we expanded the field of compressive sensing in both theoretical and practical ways. We first demonstrated the information scalability of CS. We applied CS principles to analog-to-digital conversion, showing ADC can be accomplished on structured high rate signals with sub-Nyquist sampling. We introduced a smashed filter to perform statistical classification problems with a rate of measurements that corresponds to the problem structure, rather than bandwidth. Second, we improved on previous work in distributed compressive sensing. We used graphical models to derive performance bounds on multi-sensor settings. Finally, we created a CS-based radar framework and applied it to both 1-D ranging and 2-D synthetic aperture problems.

15. SUBJECT TERMS
compressive sensing, manifolds, manifold learning, analog-to-digital conversion, random projections, distributed compressive sensing, graphical models, sparsity, radar

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	SAR	16	Richard G. Baraniuk
U	U	U			19b. TELEPONE NUMBER (*Include area code*) 713-348-5132

New Theory and Algorithms for *Compressive Sensing* Final Report

Richard G. Baraniuk

1 Introduction

In this report we begin by restating the motivation for our work, and review the project objectives. We present our results and follow each research thrust with potential future areas of work. We conclude with a list of publications supported by the grant, and a list of project personnel.

1.1 Review of motivation

Over the past several years, sensors and signal processing algorithms and hardware have been under increasing pressure to accommodate:

- ever larger and higher-dimensional data sets, including samples of wideband radio frequency (RF) signals, high-resolution images and video, volumetric data, three-dimensional (3-D) video, 4-D+ lightfields, and beyond;
- ever faster capture, sampling, and processing rates;
- ever lower power consumption; in order to permit remote, battery operation for long periods;
- networked sensing schemes for spatially distributed sources and phenomena;
- communication over ever more difficult channels; and
- radically new sensing modalities.

Fortunately, over the same time period, there has been an enormous increase in computational power and data storage thanks to Moore's Law, which provides a new angle to tackle these challenges.

We are currently on the verge of moving from a digital signal processing (DSP) paradigm, where analog signals are sampled periodically to create their digital counterparts for processing, to a computational signal processing (CSP) paradigm, where analog signals will be converted (often directly) to any of a number of intermediate representations for processing using computational and optimization techniques. At the foundation of CSP lie new uncertainty principles that generalize Heisenberg's between the time and frequency domains, the concepts of compressibility and sparsity, and the new theory of compressive sensing (CS).

The enabling idea is that natural signals and other data often contain some type of structure that makes them compressible. A compressible signal of length N can be well approximated using K real numbers, with $K \ll N$. Many audio signals, natural images, and manmade signals, for example, are compressed by a factor of 10 or more when expressed in terms of their largest Fourier or wavelet coefficients. The usual approach to acquiring a compressible signal is to take measurements in the Dirac basis and then use a nonlinear algorithm, such as a speech, MP3, JPEG, or MPEG coder, to obtain a more efficient approximation.

But this approach is not practicable if the signal is presented at a high rate (as in a radar system) or if the measurement device has limited computational resources (as in a sensor network). Fortunately, over the past two years a new theory of Compressive Sensing (CS) has emerged, in which an incoherent linear projection is used to acquire an efficient representation of a compressible signal directly using just $M \approx K \ll N$ measurements [1–6]. Interestingly, random projections play a major role. The signal is then reconstructed by solving an inverse problem either through a linear program or a greedy pursuit.

CS offers a fresh approach to framing and solving a number of timely and challenging problems in signal and image processing and imaging. In this project, we have explored its potential as a dimensionality reduction tool, as a distributed source coding system, and as a sensing framework for radar systems.

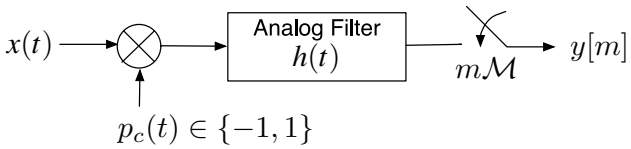

Figure 1: Pseudo-random demodulation scheme for AIC.

1.2 Review of project objectives

This project aimed at exploring the foundations and applications of CS in signal and image processing and imaging problems. Specifically, we investigated:

1. **Information scalability of CS** The CS literature has focused almost exclusively on problems in signal reconstruction, approximation, and estimation in noise. However, random projections have a long history as a dimensionality reduction tool for more general statistical modeling and classification problems [7]. We explored the information scalability of CS to a range of statistical inference tasks. In particular, we investigated how CS principles can achieve direct, high-accuracy target detection/recognition from CS measurements without reconstruction the signal/image involved and using fewer measurements. We also investigated "analog-to-information conversion," illuminating the benefits of applying CS to high-rate analog-to-digital conversion problems.

2. **Distributed sensing and encoding using CS** The CS literature has focused almost exclusively on problems involving single sensors, signals, or images. However, many important applications involve distributed networks or arrays of sensors. We developed theory and algorithms for distributed compressive sensing (DCS) that enable new signal acquisition and coding algorithms for multi-signal ensembles and sensor networks that exploit both intra- and inter-signal correlation structures. Specifically, we used graphical models to derive explicit performance bounds.

3. **CS-based radar signal processing and imaging** We investigated how CS concepts can enable new and simplified kinds of radar imaging hardware and algorithms. We formalized our approach to 1-D CS radar and expanded our existing work to a 2-D SAR CS imaging problem. We anticipate that our techniques will be particularly appropriate for inexpensive networks/arrays of radar receivers.

2 Information scalability of CS

2.1 Summary of results

Our work on information scalability was centered on two thrusts. The first was the theory and application of analog-to-information conversion. We applied CS principles to perform accurate analog-to-digital conversion on high rate signals, using a sub-Nyquist sampling rate. We developed new theory, algorithms, performance bounds, and a prototype implementation for an analog-to-information converter based on random demodulation. The architecture is particularly apropos for wideband signals that are sparse in the time-frequency plane. Our end-to-end simulations of a complete transistor-level implementation proved the concept under the effect of circuit nonidealities [8].

The second thrust was applying CS principles to detection and classification problems. Our approach was based on the generalized likelihood ratio test; in the case of image classification, it exploits the fact that a set of images of a fixed scene under varying articulation parameters forms a low-dimensional, nonlinear manifold. Exploiting recent results showing that random projections stably embed a smooth manifold in a lower-dimensional space, we developed the multiscale smashed filter as a compressive analog of the familiar matched filter classifier. In a practical target classification problem using a single-pixel camera that directly acquires compressive image projections, we achieved high classification rates using many fewer measurements than the dimensionality of the images.

2.2 Analog-to-information conversion

2.2.1 Compressive sensing background

Compressive Sensing (CS) provides a framework for acquisition of an $N \times 1$ discrete-time signal vector $\mathbf{x} = \mathbf{\Psi}\boldsymbol{\alpha}$ that is *compressible* in some *sparsity basis or frame* matrix $\mathbf{\Psi}$ (where each column is a basis or frame vector ψ_i). By

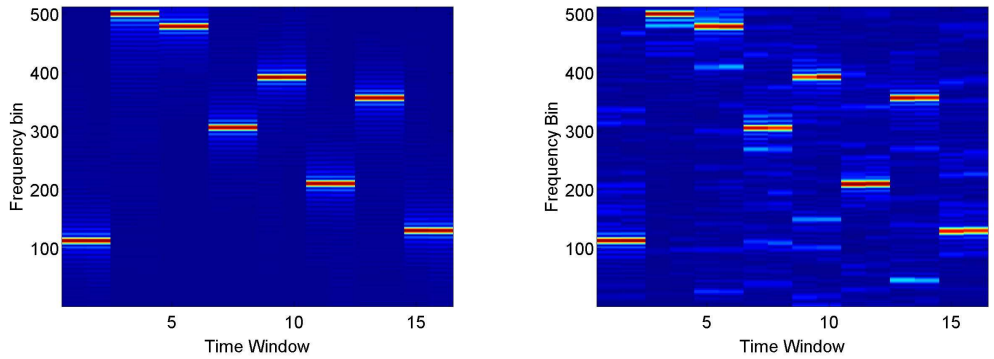

Figure 2: Comparison of Spectrograms obtained from full and CS compressed versions of a frequency hopping signal. The signal is a single side-band AM signal, whose carrier frequency changes periodically over time. (left) Spectrogram from original signal. (right) Spectrogram from CS reconstruction with measurement rate equal to 25% of Nyquist rate.

compressible we mean that the entries of $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_N]$, when sorted from largest to smallest, decay rapidly to zero; such a signal is well approximated using a K-term representation, consisting of the terms of $\boldsymbol{\alpha}$ with the K largest magnitudes while setting all the other terms to zero. Note that, by definition, signals that have only a few nonzero coefficients are compressible as well.

The CS framework [4, 6] demonstrates that a signal that is compressible in one basis $\boldsymbol{\Psi}$ can be recovered to a quality similar to that of a K-term approximation from $M = cK$ *nonadaptive* linear projections onto a second basis $\boldsymbol{\Phi}$ that is *incoherent* with the first, with c a small *overmeasuring* constant. By incoherent, we mean that the rows ϕ_j of the matrix $\boldsymbol{\Phi}$ cannot sparsely represent the elements of the sparsity-inducing basis ψ_i, and vice versa. Thus, rather than measuring the N-point signal $\mathbf{x}$ directly, we acquire the $M \ll N$ linear projections $\mathbf{y} = \boldsymbol{\Phi}\mathbf{x} + \mathbf{n} = \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\alpha} + \mathbf{n}$, where $\mathbf{n}$ represents the noise inherent to the measurement process. For brevity, we define the $M \times N$ matrix $\boldsymbol{\Theta} = \boldsymbol{\Phi}\boldsymbol{\Psi}$.

Since $M < N$, recovery of the signal $\mathbf{x}$ from the measurements $\mathbf{y}$ is ill-posed in general; however, the additional assumption of signal *compressibility* in the basis $\boldsymbol{\Psi}$ makes recovery both feasible and practical. The recovery of the set of transform coefficients $\boldsymbol{\alpha}$ can be achieved through *optimization* [9] by searching for the signal with the smallest ℓ_1 norm for the coefficient vector $\boldsymbol{\alpha}$ that agrees with the M observed measurements in $\mathbf{y}$ within the margin of error given by the magnitude of the noise $\epsilon \geq \|\mathbf{n}\|_2$:

$$\widehat{\boldsymbol{\alpha}} = \arg\min \|\boldsymbol{\alpha}\|_1 \quad \text{such that } \|\mathbf{y} - \boldsymbol{\Theta}\boldsymbol{\alpha}\|_2 \leq \epsilon \tag{1}$$

This optimization problem, also known as *Basis Pursuit with Denoising* (BPDN) [10] can be solved with traditional convex programming techniques whose computational complexities are polynomial in N. At the expense of slightly more measurements, iterative greedy algorithms like Orthogonal Matching Pursuit (OMP) [11] can also be applied to the recovery problem.

2.2.2 Real-time CS

Our signal acquisition system consists of three main components; demodulation, filtering, and uniform sampling. As seen in Figure 1, the signal is modulated by a psuedo-random maximal-length PN sequence of ± 1's. We call this the *chipping sequence* $p_c(t)$; its chipping rate, i.e. the rate of change of symbols, must be faster than the Nyquist rate for the input signal. The purpose of such modulation is to provide randomness necessary for successful CS recovery. The modulation is followed by a low-pass filter with impulse response $h(t)$. Finally, the signal is sampled at rate $\mathcal{M}$ using a traditional ADC. This system can be formulated as a CS measurement matrix as seen in [12].

2.2.3 Reconstruction for analog time-frequency sparse signals

We consider the case of wideband signals that are time-frequency sparse in the sense that at each point in time they are well-approximated by a few local sinusoids of constant frequency. As a practical example, consider sampling a frequency-hopping communications signal that consists of a sequence of windowed sinusoids with frequencies distributed between f_1 and f_2 Hz. The bandwidth of this signal is $f_2 - f_1$ Hz, which dictates sampling above the

Nyquist rate of $2(f_2 - f_1)$ Hz to avoid aliasing. We are interested in the case where $f_2 - f_1$ is very large and the signal is compressible, since the AIC will achieve much better performance than an ADC.

It is well known that signals that are localized in the time-frequency domain have compact transformations under the Gabor transform, which is defined as

$$\widehat{x}(\tau, f) = \langle x(t), \psi_{\tau,f}(t) \rangle,$$

i.e. the coefficient measures the inner product of the signal with the Gabor atoms

$$\psi_{\tau,f}(t) = g(t - \tau)e^{\pm j2\pi ft}$$

where g is a window function with $\|g\|_2 = 1$ [13]. We will leverage this compact nature during the reconstruction of the signal to obtain a representation directly in the time-frequency domain, without performing reconstruction of the original time signal.

The conventional tool for this class of signals is a *spectrogram*. A spectrogram is assembled using the magnitude of short-time Fourier transforms (STFT) that performs Fourier analysis of shifted windowed versions of the input signals to establish frequency content at local time neighborhoods. The STFT is written as

$$\alpha[l, m] = \left\langle x, \psi_{l\tau, \frac{m}{n}} \right\rangle = \int_{-\infty}^{\infty} x(t)g(t - l\tau)e^{-j2\pi mt/n}dt$$

for $l = 1, \ldots, n/\tau$ and $m = 1, \ldots, n$. This tool provides a visual representation of the Fourier spectrum of a signal over time. The spectrogram can be thought of as a uniform sampling of the coefficients of the signal under the Gabor transform. Thus, by utilizing a dictionary matrix $\mathbf{\Psi}$ consisting of a sampling of the Gabor atoms, the signal $\mathbf{x}$ can be represented using a sparse or compressible vector $\boldsymbol{\alpha}$ under the dictionary $\mathbf{\Psi}$. In this fashion, our sparse reconstruction of the signal will be obtained directly in the time frequency domain - we observe the spectrogram directly without requiring reconstruction of the original signal. An example is shown in Figure 2(a) where the spectrogram of a single sideband AM frequency hopping signal is displayed. We see that for small ranges of time, the signal is well identified by its carrier frequency, but when we consider the whole signal length there are many carriers to isolate. The spectrogram pictured in Figure 2(b) shows reconstruction of the signal from AIC measurements using a Gabor dictionary with a boxcar window. The carriers in the reconstruction are easily identified. The noise appears due to the non-sparse structure of the input signal; however, its compressibility allows us to recover the largest components.

As a bonus, when the we reconstruct the sparse representation $\boldsymbol{\alpha}$ from our measurements $\mathbf{y}$, the values in $\boldsymbol{\alpha}$ directly correspond to the coefficients in the spectrogram. This is apparent from the formulation of the Gabor atoms and the STFT. A spectrogram analysis can be immediately displayed from $\boldsymbol{\alpha}$ without final reconstruction of the signal's estimated time representation $\widehat{\mathbf{x}}$.

2.2.4 Analog-to-information system performance

In this section we wish to characterize the SNR of the AIC system using known analysis of CS performance. We present a theorem for K-sparse signals, which gives insight into the SNR behavior of the AIC system. The following definition is used in the theorem.

Definition 1 *A matrix Φ of size $M \times N$ holds the K-Restricted Isometry Property (K-RIP) with constant δ_K if for all $\mathbf{x} \in \mathbb{R}^N$ with $\|\mathbf{x}\|_0 = K$,*

$$(1 - \delta_K)\|\mathbf{x}\|_2 \leq \|\Phi\mathbf{x}\|_2 \leq (1 + \delta_K)\|\mathbf{x}\|_2.$$

Theorem 1 *Let $\mathbf{x}$ be an K-sparse signal, i.e. $\|\mathbf{x}\|_0 = K$, and let $\mathbf{y} = \Phi\mathbf{x}$ represent an AIC measurement setup, where we label reconstruction from the measurements $\mathbf{y}$ as $\mathbf{x}^\sharp$ with AIC reconstruction using BPDN. If Φ holds the K-Restricted Isometry Property (RIP) with constant δ_K and if $\delta_{3K} + 3\delta_{4K} < 2$, then the SNR of the AIC system obeys the lower bound*

$$\begin{aligned} SNR_{AIC} &= 20\log\left(\frac{\|\mathbf{x}\|_2}{\|\mathbf{x}^\sharp - \mathbf{x}\|_2}\right) \\ &\geq SNR_{system} - 20\log((1 + \delta_K)C_{1,K}) \end{aligned}$$

where SNR_{system} is the SNR of the sampling subsystem and $C_{1,K}$ is a constant depending only on K.

The condition on the RIP constants holds for random Gaussian matrices when the number of rows is large enough. The theorem is proven in [14]. This bound on the performance decay will depend on the compressibility of the signal and the class of matrix Φ applied. As an example, if a Gaussian random matrix Φ is used with a large enough row-to-column ratio, and the signal has a sparsity $K = N/10$, the loss in performance is approximately 23dB.

2.3 The smashed filter

As the second thrust of our work on information scalability, we formulated a classification algorithm that uses compressive multiscale measurements to exploit the low-dimensional manifold structure inherent in the signal classes used in target recognition applications. We learned this manifold structure from training data, which served as a sampling of points from the manifolds. Such structure allowed us both to reduce the dimensionality of the training data through random measurements, and to limit the amount of training data required to perform the classification.

2.3.1 Generalized likelihood ratio test

In our setting, we have P possible classes and we define the hypothesis $\mathcal{H}_i$ to be that the observed image $\mathbf{x} \in \mathbb{R}^N$ belongs to class $\mathcal{C}_i$ for $i = 1, \ldots, P$. For each class $\mathcal{C}_i$, an element $\mathbf{x} \in \mathcal{C}_i$ can be parameterized by a unique K-dimensional parameter vector $\boldsymbol{\Theta}_i \in \Theta_i$, i.e. $\mathbf{x} = f_i(\boldsymbol{\Theta}_i)$ for some f_i; an example parameter is the pose of the object in the scene (translation, rotation, etc.). If the mapping f_i is well-behaved, the collection of signals $\{f_i(\boldsymbol{\Theta}_i) : \boldsymbol{\Theta}_i \in \Theta_i\}$ forms a K-dimensional manifold embedded in the ambient signal space.

We will first assume that noisy measurements of $\mathbf{x}$ are taken, $\mathbf{y} = \mathbf{x} + \boldsymbol{\omega}$, giving us a distribution $p(\mathbf{y}|\boldsymbol{\Theta}_i, \mathcal{H}_i)$ for the measured signal $\mathbf{y}$ under hypothesis $\mathcal{H}_i$ and parameters $\boldsymbol{\Theta}_i$. The GLRT classifier is

$$\mathcal{C}(\mathbf{y}) = \arg\max_{i=1,\ldots,P} \; p(\mathbf{y}|\widehat{\boldsymbol{\Theta}_i}, \mathcal{H}_i), \tag{2}$$

where

$$\widehat{\boldsymbol{\Theta}_i} = \arg\max_{\theta \in \Theta_i} \; p(\mathbf{y}|\boldsymbol{\Theta}, \mathcal{H}_i) \tag{3}$$

denotes the maximum likelihood estimate (MLE) of the parameters $\boldsymbol{\Theta}_i$ under hypothesis $\mathcal{H}_i$. Under an additive white Gaussian noise (AWGN) model for $\boldsymbol{\omega}$, the likelihood for each hypothesis $\mathcal{H}_i$ becomes

$$p(\mathbf{y}|\widehat{\boldsymbol{\Theta}_i}, \mathcal{H}_i) \propto \frac{1}{\|\mathbf{y} - f_i(\widehat{\boldsymbol{\Theta}_i})\|_2^2}, \tag{4}$$

meaning that after estimates for the parameters are obtained for each class, the GLRT reduces to nearest-neighbor classification among the available hypotheses.

2.3.2 Manifold parameter estimation

In order to implement the GLRT as described above, we first need to obtain estimates of the parameter vectors $\widehat{\boldsymbol{\Theta}_i}$ from the noisy measurements $\mathbf{y}$ under each of the hypotheses. A natural approach to this problem is through nonlinear least-squares, in which we seek the value of $\boldsymbol{\Theta}_i$ that minimizes the objective function

$$D(\boldsymbol{\Theta}_i) = \|\mathbf{y} - f_i(\boldsymbol{\Theta}_i)\|_2^2. \tag{5}$$

For differentiable $D(\boldsymbol{\Theta})$, we can use Newton's method to obtain iterative estimates of the parameters as

$$\boldsymbol{\Theta}_i^n = \boldsymbol{\Theta}_i^{n-1} - [\mathbf{H}(\boldsymbol{\Theta}_i^{n-1})]^{-1} J(\boldsymbol{\Theta}_i^{n-1}) \tag{6}$$

for the n^{th} iteration, with $\mathbf{J}(\boldsymbol{\Theta}) = \vec{\nabla} D(\boldsymbol{\Theta})$ (the gradient) and $\mathbf{H}(\boldsymbol{\Theta})$ the Hessian matrix of D; with a good initial choice the algorithm converges to the correct estimate. Note that the classical matched filter is an elegant method for minimizing (5) on the manifold consisting of all possible shifts of a signal. In (6) we extend this approach to arbitrary differentiable manifolds. In essence, (6) provides a way of generalizing the classical matched filter to a richer class of manifolds, while reducing the number of samples from the manifold required during the estimation process.

However, in extending this to our compressive classification setting, we face a number of challenges. First, in general, implementing such an estimator requires complete knowledge of the function f_i or the ability to evaluate $f_i(\boldsymbol{\Theta})$ for all possible values of $\boldsymbol{\Theta}$. In some practical settings this may not be possible, but this is easily overcome since a dense sampling of the parameter space Θ_i and a nearest neighbor (NN) estimation rule can give acceptable performance, albeit with a potentially high computational cost. Potentially more challenging is that: (i) our manifolds may not be differentiable, in which case we cannot directly apply (6) [15], and (ii) it may be possible that random projections of our data could alter the manifold structure of our signals. Fortunately, we can overcome both of these challenges through the use of *multiscale* measurements.

2.3.3 Multiscale measurements for image appearance manifolds

In the case of interest—target classification—the classes $\mathcal{C}_i$ are IAMs that each correspond to different classes of targets. The parameter vector $\boldsymbol{\Theta}_i$ denotes the articulation parameters for the target, such as rotation, translation, angle of view, etc. The resulting parametric manifolds are *nonlinear*—since linear combinations of manifold elements are in general not contained in the manifold—and *non-differentiable*—due to prevalent changes in hard edges in the image view caused by rotations in and out of view, occlusions, etc. Previous research [15] has identified a multiscale structure to such manifolds that can be exploited through regularization to allow differentiability at several scales. This is achieved through the use of a nested set of regularization kernels $G_1, G_2, \ldots$ for each iteration, with the kernels becoming increasingly sharper. Thus, instead of applying Newton's method to $f(\boldsymbol{\Theta}_i)$ directly, we use an objective function for the regularized images; for the n^{th} iteration, the objective function becomes

$$D_n(\boldsymbol{\Theta}_i) = \|G_n \mathbf{y} - G_n f_i(\boldsymbol{\Theta}_i)\|_2^2, \tag{7}$$

which uses the corresponding regularization kernel.

2.3.4 Compressive measurements for smooth manifolds

It has also been shown that most of the structure of a *smooth* signal manifold is preserved under a random lower dimensional projection [16]. More specifically, for a K-dimensional manifold embedded in N-dimensional space, with high probability a random M-dimensional projection is invertible—and thus preserves the manifold structure—provided that $M > CK\log(N)$ for some constant C that depends on the smoothness of the manifold. Thus, instead of performing parameter estimation based on a direct measurement of the signal $\mathbf{y} = \mathbf{x} + \boldsymbol{\omega}$, we can choose to observe only a lower dimensional, randomly projected version $\mathbf{y} = \Phi\mathbf{x} + \boldsymbol{\omega}$, where Φ is an $M \times N$ measurement matrix with independent, randomly distributed entries. Accordingly, we update the objective function to

$$D^C(\boldsymbol{\Theta}_i) = \|\mathbf{y} - \Phi f_i(\boldsymbol{\Theta}_i)\|_2^2. \tag{8}$$

When Gaussian random measurements are used, this is equivalent to employing different colored Gaussian random measurements at each iteration; see [17] for more details. Moreover, the dimensionality reduction affords savings in computational complexity and storage requirements of the estimation and classification algorithms described earlier.

2.3.5 The multiscale smashed filter

We are now in a position to describe how we will overcome the challenges listed at the end of Section 2.3.2. In [18] we introduced the *smashed filter* as a method for classification that uses compressive measurements for classification where each class is represented by a low-dimensional manifold. This is inspired by the fact that random projections do not disturb the structure of smooth manifolds, as described above. However, as we have just observed, in our setting the manifolds might not be smooth. To address this problem, we exploit the multiscale structure of IAMs and combine the use of multiscale measurements with random projections. Thus we smooth the IAMs so that the projections preserve their geometry. This uses a measurement matrix of the form

$$\Phi = \begin{bmatrix} \Phi_1 G_1 \\ \vdots \\ \Phi_S G_S \end{bmatrix},$$

where Φ_n is an $M_n \times N$ matrix with randomly distributed entries and G_n is the regularization kernel for the n^{th} scale. The resulting measurements can be partitioned into measurements for each of the regularized versions, i.e., $\mathbf{y}_n = \Phi_n G_n \mathbf{x} + \boldsymbol{\omega}_n$, which are used on sequential iterations of Newton's method by employing the corresponding objective functions

$$D_n^C(\boldsymbol{\Theta}_i) = \|\mathbf{y}_n - \Phi_n G_n f_i(\boldsymbol{\Theta}_i)\|_2^2. \tag{9}$$

This classification algorithm, which we call the *multiscale smashed filter*, employs the compact and multiscale nature of the manifolds defined by the signal classes to estimate the signal parameters under each class hypothesis, together with the GLRT/NN classification rule from Section 2.3.1.

2.3.6 Advantages of compressive classification

In addition to the computational and storage savings achieved by compressive classification, our proposed method shares many advantages previously shown for CS reconstruction. In particular, random projections allow for *universal* estimation and classification, in the sense that random projections preserve any low-dimensional structure of a signal class with high probability. Additionally, we attain progressivity in the sense that a larger number of projections translate into higher classification rates due to increased noise tolerance.

(a) Tank (b) School Bus (c) Truck

Figure 3: *Models used for classification experiments.*

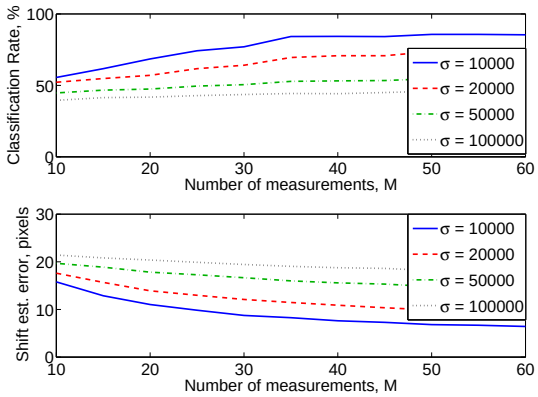

Figure 4: *Vehicle classification result from compressive imaging measurements using the smashed filter. The probability of classification and the position estimation error improve as the number of measurements increases.*

2.3.7 Experimental performance

We performed experiments to evaluate the multiscale smashed filter in a target classification setting using synthetically generated binary random measurements. In these experiments we define three classes, each for a different vehicle model: a T-72 tank, a school bus and a truck. The unknown parameter in each vehicle class is the location of the vehicle in the image, which can vary in an area of 32×32 pixels. The models used are shown in Figure 3. For each of the vehicles, multiscale measurements were taken using five different resolutions – from 8×8 and 128×128 pixels – with the same number of measurements taken at each of them.

We tested the performance of the multiscale smashed filter classifier under different levels of Gaussian noise. The measurements for each position/class combination were classified using a multiscale smashed filter trained on all other available data points. For each of the target classes, one of the sampled rotations was chosen at random as an initial estimate. The gradient of the manifold $f_i(\boldsymbol{\Theta}_i)$ was estimated using consecutive points in the manifold sampling for each parameter, including that of the current estimate. We then executed Newton's method using measurements at different resolutions at each iteration, proceeding from the coarsest to the finest scale. After the position was estimated under each hypothesis, nearest neighbor classification was performed. We repeated the experiment 10,000 times for each testing point, with randomly selected starting points each time, and we varied the number of measurements taken from 5 to 60. We also varied the power of the noise added to the measurement vector Results are shown in Figure 4, and show that due to the low-dimensional structure of the underlying IAM, very few measurements are necessary to achieve high classification rates. Additionally, the performance of the algorithm degrades gracefully as the power of the noise present increases.

2.4 Future work

The results of both of our thrusts lead to several areas of future work. Analog-to-information theory and practice can be applied to any sensing scenario in which the volume of data exceeds traditional analog-to-digital conversion abilities, or makes them cost-prohibitive. Therefore continued exploration into implementation, reconstruction technique, performance analysis, and robustness to noise would be fruitful. With the compressive classification thrust, we hope to develop more sophisticated algorithms to exploit the manifold structure to more efficiently obtain the ML estimates required by the smashed filter. For example, rather than an exhaustive nearest-neighbor search, which could be computationally prohibitive for a large training set, a greedy approach might offer similar

performance at significant computational savings; other approaches that exploit the smoothness of the manifolds could also be beneficial.

3 Distributed compressive sensing

3.1 Summary of results

The CS framework has been proposed for efficient acquisition of sparse and compressible signals through incoherent measurements. In past work, we introduced a new concept of joint sparsity of a signal ensemble and used them in demonstrating distributed CS schemes. In this project we considered joint sparsity via graphical models that link the sparse underlying coefficient vector, signal entries, and measurements. Our main results are converse and achievable bounds establishing that the number of measurements required in the noiseless measurement setting is closely related to the dimensionality of the sparse coefficient vector. Single signal and joint (single-encoder) CS are special cases of joint sparsity, and their performance limits fit into our graphical model framework for distributed (multi-encoder) CS.

3.2 Review of joint sparsity models

In this section, we generalize the notion of a signal being sparse in some basis to *joint sparsity* within a signal ensemble. We begin with basic notation. Let $\Lambda := \{1, 2, \ldots, J\}$ be the set of signal indices. Denote the signals in the ensemble by $x_j \in \mathbb{R}^N$, where $j \in \Lambda$. We use $x_j(n)$ to denote sample n in signal j, and assume for the sake of illustration that these signals are sparse in the canonical basis, i.e., $\Psi = \mathbf{I}$. The entries of the signal can take arbitrary real values, and the framework is extendable to arbitrary Ψ.

We denote by Φ_j the measurement matrix for signal j; Φ_j is $M_j \times N$ and, in general, entries of Φ_j are different for each j. Thus, $y_j = \Phi_j x_j$ consists of $M_j < N$ *random measurements* of x_j. We emphasize random Gaussian matrices Φ_j in the following, but other measurement matrices are possible. To compactly represent the signal and measurement ensembles, we define $X = [x_1^T \ \ldots \ x_J^T]^T \in \mathbb{R}^{JN}$ and $Y = [y_1^T \ \ldots \ y_J^T]^T \in \mathbb{R}^{\sum M_j}$. Finally, we also define $\Phi = \mathrm{diag}(\Phi_1, \ldots, \Phi_J)$, where diag denotes a matrix diagonal concatenation, to get $Y = \Phi X$.

3.2.1 Algebraic framework

Our framework enables analysis of a given ensemble $x_1, x_2, \ldots, x_J$ in a "jointly sparse" sense, as well as a metric for the complexities of different signal ensembles. It is based on a factored representation of the signal ensemble, and decouples location and value information. We begin by illustrating the single signal case.

Single signal case: Consider a sparse $x \in \mathbb{R}^N$ with $K < N$ nonzero entries. Alternatively, we can write $x = P\theta$, where $\theta \in \mathbb{R}^K$ contains the nonzero values of x, and P is an *identity submatrix*, i.e., P contains K columns of the $N \times N$ identity matrix $\mathbf{I}$. To model the set of all possible sparse signals, let $\mathcal{P}$ be the set of all identity submatrices of all possible sizes $N \times K'$, with $1 \leq K' \leq N$. We refer to $\mathcal{P}$ as a *sparsity model*. Given a signal x, one may consider all possible factorizations $x = P\theta$, with $P \in \mathcal{P}$. Among them, the smallest dimensionality for θ indicates the *sparsity* of x under the model $\mathcal{P}$.

Multiple signal case: For multiple signals, consider factorizations of the form $X = P\Theta$ where $X \in \mathbb{R}^{JN}$ as above, $P \in \mathbb{R}^{JN \times D}$, and $\Theta \in \mathbb{R}^D$. We refer to P and Θ as the *location matrix* and *value vector*, respectively. A *joint sparsity model* (JSM) is defined in terms of a set $\mathcal{P}$ of admissible location matrices P with varying numbers of columns. Unlike the single signal case, there are multiple choices for what matrices P belong to a joint sparsity model $\mathcal{P}$.

Minimal sparsity: For a given ensemble X, let $\mathcal{P}_F(X)$ denote the set of feasible location matrices $P \in \mathcal{P}$ for which a factorization $X = P\Theta$ exists. Among the feasible location matrices, we let $\mathcal{P}_M(X) \subseteq \mathcal{P}_F(X)$ denote the matrices P having the minimal number of columns. The number of columns D for each $P \in \mathcal{P}_M(X)$ is called the *joint sparsity level* of X under the model $\mathcal{P}$. Generally speaking, the minimal location matrices $\mathcal{P}_M(X)$ permit the most efficient factorizations of the signal ensemble; we show in Section 3.3 that these matrices dictate the number of measurements.

We restrict our attention in this paper to scenarios where each signal x_j is generated as a combination of two components: (*i*) a common component z_C, which is present in all signals, and (*ii*) an innovation component z_j, which is unique to each signal. These combine additively, giving $x_j = z_C + z_j, \quad j \in \Lambda$. However, individual components might be zero-valued in specific scenarios.

3.2.2 Example joint sparsity model: JSM-1

In the sparse common and innovations (JSM-1) model [19], the common component z_C and each innovation component z_j are *sparse* with respective sparsities K_C and K_j. Within our algebraic framework, the class of JSM-1 signals correspond to the set of all matrices

$$P = \begin{bmatrix} P_C & P_1 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ P_C & \mathbf{0} & \dots & P_J \end{bmatrix},$$

where P_C and $\{P_j\}_{j\in\Lambda}$ are arbitrary identity submatrices of sizes $N \times K_C$ and $N \times K_j$, respectively, and $\mathbf{0}$ denotes a zero matrix of appropriate size. Given $X = P\Theta$, we can partition the value vector $\Theta = [\theta_C^T\ \theta_1^T\ \theta_2^T\ \dots\ \theta_J^T]^T$, where $\theta_C \in \mathbb{R}^{K_C}$ and each $\theta_j \in \mathbb{R}^{K_j}$. When generating a signal according to this model, we have $z_C = P_C\theta_C$, $z_j = P_j\theta_j$, $j \in \Lambda$. If $P \in \mathcal{P}_M(X)$, then the joint sparsity is $D = K_C + \sum_{j\in\Lambda} K_j$.

Sparsity reduction: If a signal ensemble $X = P\Theta$, $\Theta \in \mathbb{R}^D$, were to be generated by a selection of P_C and $\{P_j\}_{j\in\Lambda}$, where all $J+1$ identity submatrices share a common column vector, then $P \notin \mathcal{P}_M(X)$. By removing the instance of this column in P_C, one obtains $Q \in \mathcal{P}$ such that there exists $\Theta' \in \mathbb{R}^{D-1}$ with $X = Q\Theta'$. We term this phenomenon *sparsity reduction*, since it reduces the effective joint sparsity of a signal ensemble.

3.3 DCS goal: Bound on measurements rates

We seek conditions on the number of measurements from each sensor that guarantee perfect recovery of X given Y. Within our algebraic framework, recovering X involves determining a value vector Θ and location matrix P such that $X = P\Theta$. Two challenges are present. First, a given measurement depends only on some of the components of Θ, and the measurement budget should be adjusted between the sensors in order to gather sufficient information on all components of Θ. Second, the decoder must identify a feasible location matrix $P \in \mathcal{P}_F(X)$ from the set $\mathcal{P}$ and the measurements Y. In this section, we develop tools to address these challenges and characterize the number of measurements needed by them.

3.3.1 Graphical model framework

We introduce a graphical representation that captures the dependencies between the measurements in Y and the value vector Θ, represented by Φ and P. Consider a feasible decomposition of X into $P \in \mathcal{P}_F(X)$ and the corresponding Θ. We define the following sets of vertices, illustrated in Figure 5(a): (*i*) the set of *value vertices* V_V has elements with indices $d \in \{1, \dots, D\}$ representing entries of the value vector $\theta(d)$; (*ii*) the set of *signal vertices* V_S has elements with indices (j, n) representing the signal entries $x_j(n)$, with $j \in \Lambda$ and $n \in \{1, \dots, N\}$; and (*iii*) the set of *measurement vertices* V_M has elements with indices (j, m) representing the measurements $y_j(m)$, with $j \in \Lambda$ and $m \in \{1, \dots, M_j\}$. The cardinalities of these sets are $|V_V| = D$, $|V_S| = JN$ and $|V_M| = \sum_{j\in\Lambda} M_j$.

Let P be partitioned into *location submatrices* P^j, $j \in \Lambda$, so that $x_j = P^j\Theta$; here P^j is the restriction of P to the rows that generate the signal x_j. We then define the bipartite graph $G = (V_S, V_V, E)$, determined by P, where there exists an edge connecting (j, n) and d if and only if $P^j(n, d) \neq 0$.

A similar bipartite graph $G' = (V_M, V_S, E')$, illustrated in Figure 5(a), connects between the measurement vertices $\{(j, m)\}$ and the signal vertices $\{(j, n)\}$; there exists an edge in G' connecting $(j, n) \in V_S$ and $(j, m) \in V_M$ if $\Phi_j(m, n) \neq 0$. When the measurements matrices Φ_j are dense, which occurs with probability one for i.i.d. Gaussian random matrices, the vertices corresponding to entries of a given signal x_j in V_S are all connected to all vertices corresponding to the measurements y_j in V_V. Figure 5 shows an example for dense measurement matrices: each measurement vertex $(j, \cdot)$ is connected to each signal vertex $(j, \cdot)$.

The graphs G and G' can be merged into $\widehat{G} = (V_M, V_V, \widehat{E})$ that relates entries of the value vector to measurements. Figure 5(b) shows the example composition of the previous two bipartite graphs. $\widehat{G}$ is used to recover Θ from the measurement ensemble Y when P is known.

3.3.2 Quantifying dependencies and redundancies

We now define the subset of the value vector entries that is measured exclusively by a subset Γ of the sensors in the ensemble; the cardinality of this set will help determine the number of measurements the sensors in Γ should perform. We denote by $E(V)$ the neighbors of a set of vertices V through the edge set E.

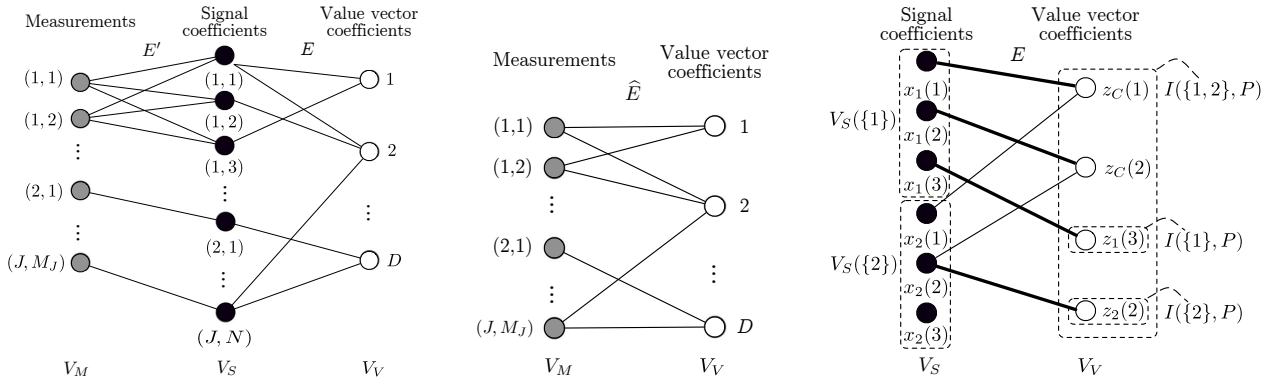

Figure 5: *Bipartite graphs for distributed compressed sensing. (a) $G = (V_S, V_V, E)$ connects the entries of each signal with the value vector coefficients they depend on; $G' = (V_M, V_S, E')$ connects the measurements at each sensor with observed signal entries. The matrix Φ is a dense Gaussian random matrix, as shown in the graph. (b) $\widehat{G} = (V_M, V_V, \widehat{E})$ is the composition of G and G', and relates between value vector coefficients and measurements. (c) Sets of exclusive indices for our example.*

Definition 2 *Let $G = (V_S, V_V, E)$ be the bipartite graph determined by P, let $\Gamma \subseteq \Lambda$, and let $V_S(\Gamma)$ be the set of vertices $V_S(\Gamma) = \{(j,n) \in V_S : j \in \Gamma, n \in \{1,\dots,N\}\}$. We define the* set of exclusive indices *for Γ given P, denoted $I(\Gamma, P)$, as the largest subset of $\{1,\dots,D\}$ such that $E(I(\Gamma,P)) \subseteq V_S(\Gamma)$.*

$I(\Gamma, P)$ is significant in our distributed measurement setting, because it contains the coefficients of θ that only affect the signals in the set Γ and, therefore, can only be measured by those sensors. Figure 5(c) shows an example setting of two signals of length $N = 3$ generated by a matrix P from the JSM-1 model, with the sets $I(\{1\}, P)$ and $I(\{2\}, P)$ defined as the vertices in V_V that connect exclusively with $V_S(\{1\})$ and $V_S(\{2\})$, respectively.

Overlaps: When overlaps between common and innovation components are present in a signal, we cannot recover the overlapped portions of both components from the measurements of this signal alone; we need to recover the common component's coefficients using measurements of other signals that do not feature the same overlap. Furthermore, these coefficients of the value vector are not included in $I(\Gamma, P)$. We thus quantify the size of the overlap for all subsets of signals $\Gamma \subset \Lambda$ under a feasible representation given by P and Θ.

Definition 3 *The* overlap size *for the set of signals $\Gamma \subset \Lambda$, denoted $K_{C,\Gamma}$, is the number of indices in which there is overlap between the common and the innovation component supports at the signals $j \notin \Gamma$; more formally,*

$$K_{C,\Gamma}(P) = |\{n \in \{1,\dots,N\} : z_C(n) \neq 0,\ \forall j \notin \Gamma, z_j(n) \neq 0\}|.$$

For the entire set of signals, the overlap size $K_{C,\Lambda} = 0$.

For $\Gamma \neq \Lambda$, $K_{C,\Gamma}(P)$ provides a penalty term due to the need for recovery of common component coefficients that are overlapped by innovations in all other signals $j \notin \Gamma$. The definition of $K_{C,\Lambda}$ accounts for the fact that all the coefficients of Θ are included in $I(\Lambda, P)$.

3.3.3 Main results

Converse and achievable bounds on the number of measurements necessary for recovery are given below.

Theorem 2 (Achievable, known P) *Assume that a signal ensemble X is obtained from a common/innovation component JSM $\mathcal{P}$. Let $\{M_j\}_{j\in\Lambda}$ be a measurement tuple. Suppose there exists a full rank location matrix $P \in \mathcal{P}_F(X)$ such that*

$$\sum_{j\in\Gamma} M_j \geq |I(\Gamma, P)| + K_{C,\Gamma}(P) \tag{10}$$

for all $\Gamma \subseteq \Lambda$. If the Φ_j are random matrices having M_j rows of i.i.d. Gaussian entries for each $j \in \Lambda$, and if $Y = \Phi X$, then with probability one over Φ, there is a unique solution $\widehat{\Theta}$ to the system of equations $Y = \Phi P \widehat{\Theta}$, and hence the signal ensemble X can be uniquely reconstructed as $X = P\widehat{\Theta}$.

Theorem 3 (Achievable, unknown P) *Assume that a signal ensemble X is obtained from a common/innovation component JSM $\mathcal{P}$, and let Φ_j be random matrices having M_j rows of i.i.d. Gaussian entries for each $j \in \Lambda$. If there exists a location matrix $P^* \in \mathcal{P}_F(X)$ such that*

$$\sum_{j \in \Gamma} M_j \geq |I(\Gamma, P^*)| + K_{C,\Gamma}(P^*) + |\Gamma| \tag{11}$$

for all $\Gamma \subseteq \Lambda$, then X can be uniquely recovered from Y with probability one over Φ.

Theorem 4 (Converse) *Assume that a signal ensemble X is obtained from a common/innovation component JSM $\mathcal{P}$. Let $\{M_j\}_{j \in \Lambda}$ be a measurement tuple. Suppose there exists a full rank location matrix $P \in \mathcal{P}_F(X)$ such that*

$$\sum_{j \in \Gamma} M_j < |I(\Gamma, P)| + K_{C,\Gamma}(P) \tag{12}$$

for some $\Gamma \subseteq \Lambda$. Let Φ_j be any set of measurement matrices having M_j rows for each $j \in \Lambda$, and let $Y = \Phi X$. Then there exists a solution $\widehat{\Theta}$ such that $Y = \Phi P \widehat{\Theta}$ but $\widehat{X} := P\widehat{\Theta} \neq X$.

The number of measurements needed for recovery depends on the number of value vector coefficients that are observed only by the sensors in Γ. The identification of a feasible location matrix P causes the 2 measurement-per-sensor gap between the converse and achievable bounds (11-12). The algorithm used in Theorem 3 essentially performs an ℓ_0 minimization to acquire Θ, where the correct P is identified using an additional cross-validation step.

Discussion: The theorems can also be applied to the single sensor and joint measurement settings. In the single signal setting, we will have $x = P\theta$ with $\theta \in \mathbb{R}^K$, and $\Lambda = \{1\}$; the theorem provides the requirement $M \geq K + 1$, which matches the existing requirements for reconstruction.

The joint measurement setting is equivalent to the single signal setting with a dense measurement matrix, as all measurements are dependent on all signal entries. In this case, however, the distribution of the measurements among the available sensors is irrelevant. Therefore, we only obtain a condition on the total number of measurements obtained by the group of sensors as $\sum_{j \in \{1,\ldots,N\}} M_j \geq D + 1$.

3.4 Future Work

The use of graphical models allowed us to derive bounds on the number of measurements necessary to recover a signal in a multi-sensor compressive sensing setting. It follows naturally to apply the insights our graphical model framework gives us to a variety of sensing scenarios. The defined recovery technique requires an ℓ_0 minimization, so investigation of a convex optimization method and measurement rates associated with it is also a direction worth pursuing. Finally—and our graphical model approach is a step in the right direction—we still hope to bring together an over-arching theory of rate-distortion analysis to distributed compressive sensing.

4 Compressive sensing radar

4.1 Summary of results

We took the principles of CS and applied them to both 1-D ranging radar and 2-D imaging radar. We demonstrated that CS has the potential to make two significant improvements to radar systems by (i) eliminating the need for the pulse compression matched filter at the receiver, and (ii) reducing the required receiver analog-to-digital conversion bandwidth so that it need operate only at the radar reflectivity's potentially low "information rate" rather than at its potentially high Nyquist rate. These ideas could enable the design of new, simplified radar systems, shifting the emphasis from expensive receiver hardware to smart signal recovery algorithms. We formalized our approach and used it to accurately recover a 1-D ranging problem using only 50 percent of the measurements the Nyquist rate alone would dictate, and recover a 2-D SAR test image using only 25 percent of the measurements that would have been needed.

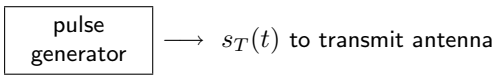

Figure 6: *Prototypical radar transmitter.*

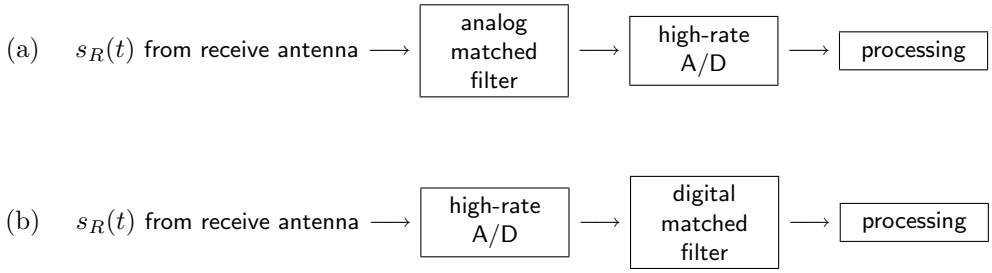

Figure 7: *Prototypical digital radar receivers for the transmitter in Fig. 6 perform matched filtering either in the (a) analog or (b) digital domain.*

4.2 CS-based radar

In order to illustrate our CS-based radar concept, consider a simplified 1D range imaging model of a target described by $u(r)$ with range variable r. If we let the transmitted radar pulse $s_T(t)$ interact with the target by means of a linear convolution [20], then the received radar signal $s_R(t)$ is given by

$$s_R(t) = A \int s_T(t-\tau)\, u(\tau)\, d\tau, \tag{13}$$

where we have converted the range variable r to time t using $t = \frac{2r}{c}$, with c the propagation velocity of light, and where A represents attenuation due to propagation and reflection. If the transmitted signal has the property that $s_T(t) * s_T(-t) \approx \delta(t)$ (which is true for PN and chirp signals), then a band-limited measurement of the radar reflectivity $u(t)$ can be obtained by pulse compression, that is, by correlating $s_R(t)$ with $s_T(t)$ in a matched filter (recall Fig. 7) [20]. A/D conversion occurs either before or after the matched filtering, resulting in N Nyquist-rate samples.

Our CS-based radar approach is based on two key observations. First, the target reflectivity functions $u(t)$ that we wish to obtain through the radar process are often *sparse* or *compressible* in some basis. For example, a set of K point targets corresponds to a sparse sum of delta functions as in $u(t) = \sum_{i=1}^{K} a_i\, \delta(t - \kappa_i)$; smooth targets are sparse in the Fourier or wavelet domain; and range-Doppler reflectivities are often sparse in the joint time-frequency (or ambiguity) domain [20]. Such target reflectivity functions $u(t)$ are good candidates for acquisition via CS techniques.

Second, time-translated and frequency-modulated versions of the PN or chirp signals transmitted as radar waveforms $s_T(t)$ form a dictionary (the extension of a basis or frame) that is *incoherent* with the time, frequency, and time-frequency bases that sparsify or compress the above mentioned classes of target reflectivity functions $u(t)$ [21]. This means that PN or chirp signals are good candidates for the rows of a CS acquisition matrix Φ as a "random filter" where:

$$y(m) = \sum_{n=1}^{N} p(Dm - n)\, x(n) \tag{14}$$

for $m = 1, \ldots, M$.

By combining these observations we can both *eliminate the matched filter in the radar receiver* and *lower the receiver A/D converter bandwidth* using CS principles. Consider a new design for a radar system that consists of the following components. The transmitter is the same as in a classical radar; the transmit antenna emits a PN or chirp signal $s_T(t)$ (recall Fig. 6). However, the receiver does not consist of a matched filter and high-rate A/D converter but rather only a low-rate A/D converter that operates not at the Nyquist rate but at a rate proportional to the target reflectivity's compressibility (see Fig. 8).

We make the connection explicit for a PN-based CS radar with a simple sampling model. Consider a target reflectivity generated from N Nyquist-rate samples $x(n)$ via $u(t) = x(\lceil t/\Delta \rceil)$, $n = 1, \ldots, N$, on the time interval of interest $0 \le t < N\Delta$. The radar transmits a PN signal generated from a length-N random Bernoulli ± 1 vector

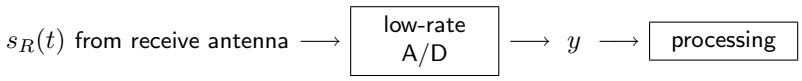

Figure 8: *Compressive radar receiver for the transmitter in Fig. 6 performs neither matched filtering nor high-rate analog-to-digital conversion.*

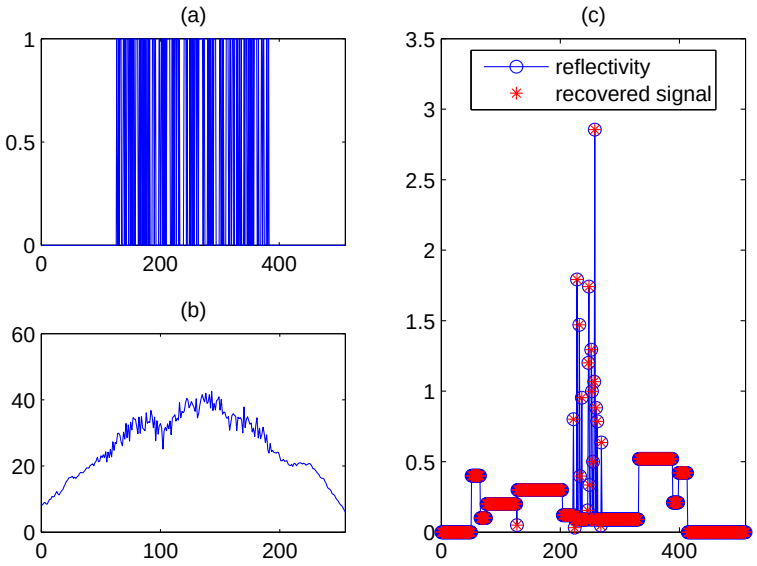

Figure 9: *CS radar example. (a) Transmitted PN pulse $s_T(t)$, (b) low-rate measurement y, and (c) true and recovered reflectivity profiles $u(t)$.*

$p(n)$ via $s_T(t) = p(\lceil t/\Delta \rceil)$. The received radar signal $s_R(t)$ is given by (13); we sample it not every Δ seconds but rather every $D\Delta$ seconds, where $D = \lfloor N/M \rfloor$ and $M < N$, to obtain the M samples, $m = 1, \ldots, M$,

$$\begin{aligned} y(m) &= s_R(t)|_{t=mD\Delta} \\ &= A \int_0^{N\Delta} s_T(mD\Delta - \tau)\, u(\tau)\, d\tau \\ &= A \sum_{n=1}^{N} p(mD-n) \int_{(n-1)\Delta}^{n\Delta} u(\tau)\, d\tau \\ &= A \sum_{n=1}^{N} p(mD-n)\, x(n), \end{aligned} \tag{15}$$

which are precisely a scaled version of (14). In words, a PN sequence radar implements a random filter in the sense of [21], and hence the low-rate samples y contain sufficient information to reconstruct the signal x corresponding to the Nyquist-rate samples of the reflectivity $u(t)$ via linear programming or a greedy algorithm. Chirp pulses yield similar results.

Figure 9 illustrates the scheme in action. A radar reflectivity profile is probed with a PN pulse sequence, measured at one-half the Nyquist sampling rate, and subsequently recovered exactly using an OMP greedy algorithm and a sparsity frame Ψ combining delta spikes and Haar wavelets.

Additional gains can be expected for 2D CS radar imaging. We illustrate this with a simple simulation of SAR data acquisition and imaging. Figure 10(a) shows the reflectivity function that is to be recovered from the SAR data. We simulated a SAR data acquisition using the method described in [22]. Figure 10(b) shows the result of a 2D CS implementation with four times undersampling, which gives an exact recovery of the reflectivity function. The traditional SAR image (Fig. 10(c)) shows artifacts of the limited aperture of the imaging operator, which are absent in the CS image. The result is similar to what is obtained with the feature-enhanced imaging approach of [23]. However, the CS-based approach has some advantages, such as an almost infinite number of sparse representations

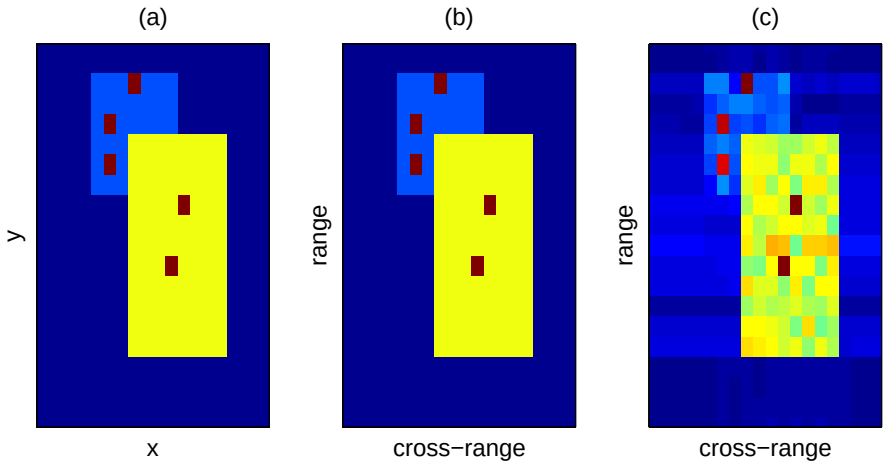

Figure 10: *CS synthetic aperture radar (SAR) example. (a) 2D reflectivity, (b) CS SAR image, and (c) traditional SAR image.*

to choose from as well as more efficient signal recovery algorithms [24].

4.3 Future work

The initial success of our approach leads us to believe that CS principles can be applied beyond the area of complete signal recovery. We showed that the information scalability of CS allows for a much wider range of statistical inference tasks. Detection, classification, and recognition would all be useful applications for Radar. The fact they require even fewer measurements than for complete reconstruction is another benefit.

Though we showed the benefits of CS-based radar, there are a number of challenges to be overcome before an actual CS-based radar system will become a reality. First, the target reflectivity being probed must be compressible in some basis, frame, or dictionary. Second, the signal recovery algorithms must be able to handle real-world radar acquisition scenarios with sufficient computational efficiency and robust performance for noisy data. Third, there is a subtle tradeoff to optimize between the reduction in sampling rate $\lfloor N/M \rfloor$ and the dynamic range of the resulting CS system [8]. These are areas of active research for both our team and the broader CS community. In particular, there could be links with recent work on finite rate of innovation sampling for ultrawideband communication systems [25].

5 Publications supported by this grant

R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *IEEE Radar Conference*, April 2007.

M. Davenport, M. Duarte, M. Wakin, J. Laska, D. Takhar, K. Kelly, and R. Baraniuk, "The smashed filter for compressive classification and target recognition," in *Computational Imaging V at SPIE Electronic Imaging*, 2007.

M. Duarte, M. Davenport, M. Wakin, J. Laska, D. Takhar, K. Kelly, and R. Baraniuk, "Multiscale random projections for compressive classification," in *IEEE Conference on Image Processing*, 2007.

M. F. Duarte, M. B. Wakin, D. Baron, and R. G. Baraniuk, "Universal distributed sensing via random projections," in *International Conference on Information Processing in Sensor Networks*, 2006.

M. Duarte, S. Sarvotham, D. Baron, M. Wakin, and R. Baraniuk, "Performance limits for jointly sparse signals via graphical models," in *Sensor, Signal and Info. Proc. Workshop*, May 2008.

J. Laska, S. Kirolos, M. Duarte, T. Ragheb, R. Baraniuk, and Y. Massoud, "Theory and implementation of an analog-to-information converter using random demodulation," in *IEEE Int. Symp. on Circuits and Systems*, 2007.

C. Rozell, D. Johnson, R. Baraniuk, and B. Olshausen, "Locally competitive algorithms for sparse approximation," in *IEEE Conference on Image Processing*, 2007.

C. Rozell, D. Johnson, R. Baraniuk, and B. Olshausen, "Sparse coding via thresholding and local competition in neural circuits," *Neural Computation*, vol. 20, pp. 2526–2563.

S. Sarvotham, D. Baron, and R. Baraniuk, "Measurements vs. bits: Compressed sensing meets information theory," in *Allerton Conference on Communication, Control, and Computing*, 2006.

M. B. Wakin and R. G. Baraniuk, "Random projections of smooth manifolds," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, May 2006.

6 Professional personnel

Principle Investigator
Richard G. Baraniuk

Research Assistant and Professional Staff (part of grant duration)
Christopher Rozell

Postdoctoral Research Associate
Petros Boufounos

Graduate Student Research Assistants
Mark Davenport, Marco Duarte, Chinmay Hegde, Jason Laska, Matthew Moravec, Shriram Sarvotham

References

[1] E. Candès and T. Tao, "Near optimal signal recovery from random projections and universal encoding strategies," *IEEE Trans. Info. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[2] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Info. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[3] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. on Information Theory*, vol. 51, pp. 4203–4215, 2005.

[4] E. Candès and J. Romberg, "Quantitative robust uncertainty principles and optimally sparse decompositions," *Found. of Comp. Math.*, vol. 6, pp. 227–254, Apr. 2006.

[5] E. Candès and J. Romberg, "Practical signal recovery from random projections," *submitted to IEEE Trans. Signal Proc.*, January 2005.

[6] D. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[7] S. Dasgupta and A. Gupta, "An elementary proof of the johnson-lindenstrauss lemma," tech. rep., 1999.

[8] J. Laska, S. Kirolos, M. Duarte, T. Ragheb, R. Baraniuk, and Y. Massoud, "Theory and implementation of an analog-to-information converter using random demodulation," in *IEEE Int. Symp. on Circuits and Systems*, 2007.

[9] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, pp. 1207–1223, Aug. 2006.

[10] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.

[11] J. Tropp and A. C. Gilbert, "Signal recovery from partial information via Orthogonal Matching Pursuit," 2005. Preprint.

[12] S. Kirolos, J. Laska, M. Wakin, M. Duarte, D. Baron, T. Ragheb, Y. Massoud, and R. Baraniuk, "Analog-to-information conversion via random demodulation," in *Proc. of the IEEE Dallas Circuits and Systems Workshop (DCAS)*, 2006.

[13] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego: Academic Press, second ed., 1999.

[14] M. F. Duarte, J. Laska, and R. G. Baraniuk, "Theoretical bounds for signal-to-noise ratio in analog to information conversion systems," Tech. Rep. TREE-0608, Rice University ECE Department, Houston, TX, Sept. 2006.

[15] M. B. Wakin and R. G. Baraniuk, "High-resolution navigation on non-differentiable image manifolds," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 5, (Philadelphia, PA), pp. V–1073–1076, 2005.

[16] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," 2006. Preprint.

[17] Z. Wang, G. R. Arce, and J. L. Paredes, "Colored projections for compressed sensing," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, (Honolulu, HI), 2007.

[18] M. A. Davenport, M. F. Duarte, D. Takhar, J. N. Laska, K. K. Kelly, and R. G. Baraniuk, "The smashed filter for compressive classification and target recognition," in *Proc. IS&T/SPIE Symposium on Electronic Imaging: Computational Imaging*, (San Jose, CA), Jan. 2007.

[19] D. Baron, M. Duarte, S. Sarvotham, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing," Tech. Rep. TREE0612, Rice University, Houston, TX, Nov. 2006. Available at `http://dsp.rice.edu/cs/`.

[20] M. Skolnik, *Radar Handbook*. New York, NY, USA: McGraw Hill, 1970.

[21] J. Tropp, M. Wakin, M. Duarte, D. Baron, and R. G. Baraniuk, "Random filters for compressive sampling and reconstruction," *Proc. IEEE ICASSP*, 2005.

[22] M. Cetin, *Feature-Enhanced Synthetic Aperture Radar Imaging*. Ph.D. Thesis Boston University, College of Engineering, 2001.

[23] M. Cetin and W. C. Karl, "Feature-enhanced synthetic aperture radar image formation based on non-quadratic regularization," *IEEE Trans. Image Processing*, vol. 10, pp. 623–631, 2001.

[24] Z. Mou-yan and R. Unbehausen, "Methods for reconstruction of 2-D sequences from Fourier transform magnitude," *IEEE Transactions on Image Processing*, vol. 6, pp. 222–233, 1997.

[25] I. Maravic, J. Kusuma, and M. Vetterli, "Low-sampling rate UWB channel characterization and synchronization," *J. Comm. and Networks*, vol. 5, Dec. 2003.